DOCUMENT RESUME

ED 292 810 TM 011 080

AUTHOR Tollefson, Nona; Tripp, Alice
TITLE The Effect of an Exclusion Item Format on Item Difficulty and Item Discrimination.
PUB DATE [86]
NOTE 15p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; *Difficulty Level; *Discriminant Analysis; Graduate Students; Higher Education; *Item Analysis; *Multiple Choice Tests; Test Construction; *Test Format; Test Items

ABSTRACT

The item difficulty and item discrimination of three multiple-choice item formats were compared in experimental and non-experimental settings. In the experimental study, 104 graduate students were randomly assigned to complete one of three forms of a multiple-choice test: (1) a complex alternative ("none of the above") as the correct answer; (2) a complex alternative as a foil; and (3) a one-correct answer format. Statistical concepts were then content tested. Analysis of variance showed significant mean differences in the item discrimination indices for the three item formats. Items having "none of the above" as the correct answer had a significantly higher mean discrimination index than did the other two item formats. There was no significant difference in the mean item difficulties for the three item formats. The non-experimental study compared the item characteristics of multiple-choice items administered as a final examination in a first-year graduate statistics course, using 44 graduate students. A significant difference in mean item difficulty was observed; the mean difference in item discrimination values was not significant. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED292810

The Effect of an Exclusion Item Format on Item Difficulty and Item Discrimination

Nona Tollefson
Professor
Educational Psychology and Research
The University of Kansas

Alice Tripp
Associate Professor
School of Nursing
Graceland College

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

NONA TOLLEFSON

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 011 080

BEST COPY AVAILABLE

The Effect of an Exclusion Item Format
on Item Difficulty and Item Discrimination

ABSTRACT

The item difficulty and item discrimination of three multiple-choice item formats were compared in experimental and non-experimental settings. In the experimental study, 104 graduate students were randomly assigned to complete one of three forms of a multiple-choice test: a complex alternative ("none of the above") as the correct answer; a complex alternative as a foil, and a one-correct answer format. Statistical concepts were the content tested. Analysis of variance showed significant mean differences in the item discrimination indices for the three item formats. Items having "none of the above" as the correct answer had a significantly higher mean discrimination index than did the other two item formats. There was no significant difference in the mean item difficulties for the three item formats. In the non-experimental study, a significant difference in mean item difficulty was observed; the mean difference in item discrimination values was not significant.

Controversy surrounds the use of complex response alternatives in multiple-choice items. Some measurement experts recommend the use of a "one-right answer" multiple-choice format (Bennion, 1977; Brown, 1976; Kryspin & Feldhusen, 1974; Mehren & Lehmann, 1978; Scannell & Tracy, 1975). Others recommend the use of complex response alternatives (Bloom, Madaus, & Hastings, 1981; Ebel, 1979).

At the present time, literature describing properties (item difficulty and item discrimination) of complex response alternatives is contradictory, and textbook writers give anecdotal reasons rather than empirical data for their recommendations to use or not to use them. A brief review of literature will illustrate the contradictory results characteristic of research on this topic.

Wesman and Bennet (1946) reported no effect on item difficulty associated with use of a "none of the above" alternative on a nursing test. Boynton (1950), Rimland (1960), and Williamson and Hopkins (1967) all reported that using "none of the above" increased item difficulty. Boynton used the exclusion option on a stenographer-typist examination. Rimland and Williamson and Hopkins used arithmetic items as their test content; however, they provided no information about the validity or reliability of the tests they used.

Studies of the effect on item difficulty and item discrimination of complex alternatives of the "all of the above" and "A and B are correct" form have also yielded contradictory results. Hughes and Trible (1965) compared a one-correct answer format to "both 1 and 2 are correct," "all of the above" and "none of the above" formats. These researchers found that the mean score for the test form including complex alternatives was sig-

nificantly lower than was the mean score for the test form including items having one-correct answer.

Mueller (1975) compared conventional multiple-choice items to three forms of complex alternatives: "all of the above," "none of the above," and combinations of alternatives such as "A and B are correct." The items were included on a real estate examination. Mueller concluded that the format using the combination of alternatives was the most difficult.

Rossie, McCrady, and Paoline (1978) compared the item discrimination of one-correct response items and matched sets of items of the "A and B, but not C" form. These authors found no significant difference in the average discrimination index of the two test forms.

Weiten (1982) compared single- and double-option multiple-choice items and found that double-option items were more difficult and less discriminating than single-option items were. Double-option items also had a lower internal consistency reliability than did single-option items. Tripp and Tollefson (1985) compared item difficulty and item discrimination of conventional multiple-choice items and parallel K-type items across six levels of the cognitive domain. The test measured nursing content. Analyses of mean item difficulty and discrimination indexes indicated that complex multiple-choice items were not more difficult nor more discriminating than conventional multiple-choice items were.

Purpose of the Study

This study compared the item difficulty and item discrimination of sets of multiple-choice items having one-correct answer and a complex alternative ("none of the above") as the correct answer. Two studies were conducted. The first study compared experimental sets of multiple-choice items having one-correct answer to parallel items having "none of the

above" as the correct answer and as a foil. The second study compared non-experimental sets of multiple-choice items having one-correct answer to items having a complex alternative as the correct answer.

Study I

Subjects

Subjects for this study were graduate students enrolled in a beginning statistics class. Four sections (n=104) of the course participated in the study. The sections were taught by two instructors, and each instructor taught two sections. The sample included 68 females and 36 males.

Instrument

The test instrument was the multiple-choice section of the final examination in the statistics course. Thirty-six multiple-choice items were randomly selected from a pool of 250 items. The item-pool had been developed by four professors, and all items had been field-tested prior to their use in the study.

From the set of 36 multiple-choice items, 10 items were randomly selected and included in the experimental set of items. Seven items had verbal response options; three items required numerical calculations and had numerical response options. Figure 1 includes examples of the types of items used in the study. Three forms of each item were developed. One form used a "one-correct answer" format, a second form used "none-of-the above" as a foil; and a third form had "none of the above" as the correct answer. Items were randomly assigned to one of the three forms of the test.

Insert Figure 1 here

In addition, four items used "the correct answer is not given" as an option. This option is commonly found in multiple-choice items measuring quantitative skills. The same set of items was included on all forms of the test.

Test forms were randomly assigned to students at the time of the final examination. Form A was completed by 38 students. It had a mean of 25.7, a standard deviation of 3.37, and internal consistency reliability of .73. Form B was completed by 34 students. The mean was 26.5, the standard deviation was 3.56, and the internal consistency reliability was .58. Form C, completed by 32 students, had a mean of 27.9, a standard deviation of 3.36, and an internal consistency reliability of .55.

Data Analysis and Results

Item difficulty (the proportion of persons answering the item correctly) and item discrimination (the point-biserial correlation between the item and the total test score) indices were computed for all items on the final examination. Items were then divided into four sets: a) the set having "none of the above" as the correct answer (E-1), b) the set having "none of the above" as a foil (E-2), c) the set using a "one-correct answer" format (E-3), and d) the set using "the correct answer is not given" format (C-4). One-way analysis of variance was used to test for over-all significant differences in the mean item difficulty and discrimination indexes for the experimental sets of items.

Table 1 reports means and standard deviations for item difficulty and item discrimination for each set of items. Item sets having "none of the above" (E-1) or "no correct answer given" as the correct answer (C-1) were the most difficult and most discriminatory. However, one-way analysis of variance yielded no significant mean differences in item difficulty values

for these sets of items (F (2,27) = 1.18, $p > .05$); significant mean differences in item discrimination values were found (F (2,27) = 4.40, $p < .05$). Items having "none of the above" as the correct answer had a significantly higher mean discrimination index than did items having "none of the above" as a foil or items having one-correct answer. There was no statistically mean significant difference in item discrimination indexes for items having "none of the above" as a foil and one-correct response.

Insert Table 1 Here

Study II

Subjects

A second study compared the item characteristics of multiple-choice items administered as a final examination in a first-year graduate statistics course. Forty-four graduate students completed the examination.

Instrument

The examination included 42 multiple-choice items. "None of the above" was the correct response for seven items; "none of the above" was a foil for 10 items; and "a and b are correct" was a foil for five items. The remaining 20 items had a one-correct answer format. The final examination had a mean of 28.9 and a standard deviation of 8.64. The Kuder-Richardson internal consistency reliability was .91.

Data Analysis and Results

Means and standard deviations were computed for item difficulty and item discrimination for items having one-correct answer and items having a complex alternative as the correct answer. Table 2 reports these means and

standard deviations. A t test yielded a significant difference in mean item difficulty for the two item sets ($t(40) = 2.17$ $p < .05$). The mean item difficulty for items with one-correct answer was significantly higher than was the mean item difficulty for items with a complex alternative as the correct answer.

Insert Table 2 here

Point-biserial correlations were used as measures of item discrimination and were transformed to Fisher's z' values to normalize the distribution of item discrimination values. A t test was used to test the hypothesis of no difference in mean discrimination z-values for items having a one-correct answer format and for items having a complex-alternative as the correct answer. This analysis yielded nonsignificant results ($t(40) = .43$ ns).

DISCUSSION

This study investigated characteristics of multiple-choice items using a complex alternative and a one-correct answer format. In the experimental study (Study I), an exclusion alternative was used as the correct answer and as a foil, and these alternatives were compared to each other and to the one-correct answer format. Items with an exclusion alternative as the correct response were consistently more difficult than were the other two item formats. However, differences in item difficulty, while consistent across four comparisons, were not large enough to be statistically significant. The non-experimental study (Study II) also found that items having a complex alternative as the correct answer were more difficult than were

items having one-correct answer. In the second study, the difference in mean item difficulty was statistically significant.

In Study I, when "none of the above" was the right answer, the mean item discrimination index was significantly higher than when "none of the above" was a foil or when a one-correct answer format was used. No significant differences in mean item discrimination values was observed in Study II. It has been argued that complex alternatives such as "none of the above" require more knowledge of content and make finer discrimination among test-takers than does a one-correct answer format. This study provided limited empirical support for this argument.

The use of "none of the above" or "no correct answer given" as the correct answer is common in quantitative tests. This study found items of this type were more difficult and more discriminating than items using a one-correct answer format. It has been argued that knowing you have solved a quantitative problem correctly and that the correct answer is not given requires a higher level of knowledge and/or comprehension than trying to solve a problem several ways in the hope of "finding one of the listed response options." This study offers limited support for this point of view.

The study, while limited in scope, suggests the need for further study of the properties of multiple-choice items using complex alternatives. Understanding the factors that influence students' responses to multiple-choice items is important both theoretically and practically. Multiple-choice items are widely used in achievement testing. Any information that increases test constructors' knowledge of the strengths and weaknesses of the various forms of multiple-choice items will add to test developers' skills in measuring student achievement.

References

Bennion, D. H. (1977). Assessing student learning; Behavioral objectives and evaluation procedures for the affective, cognitive and psychomotor domains. Dubuque, Iowa: Kendall-Hunt Publishing Company.

Bloom, B. S., Madaus, G. F., & Hastings, J. T. (1981). Evaluation to improve learning. New York: McGraw-Hill.

Boynton, M. (1950). Inclusion of 'none of these' makes spelling items more difficult. Educational and Psychological Measurement, 10, 431-432.

Brown, F. A. (1976). Principles of educational and psychological testing (2nd ed.). New York: Holt, Rinehart & Winston.

Ebel, R. L. (1979). Essentials of educational measurement (3rd ed). Englewood Cliffs, New Jersey: Prentice Hall.

Kryspin, W. J., & Feldhusen, J. F. (1974). Developing classroom tests: A guide for writing and evaluating test items. Minneapolis, Minnesota: Burgess Publishing Co.

Mehren, W. A., & Lehmann, I. J. (1978). Measurement and evaluation in education and psychology (2nd ed.). New York: Holt, Rinehart & Winston.

Mueller, D. J. (1975). An assessment of the effectiveness of complex alternatives in multiple-choice achievement test items. Educational and Psychological Measurement, 35, 135-141.

Rimland, B. (1960). The effect of varying time limits and use of 'right answer not given' in experimental forms of the U.S. Navy Arithmetic Test. Educational and Psychological Measurement, 20, 533-539.

Scannell, D. P., & Tracy, D. B. (1975). Testing and measurement in the classroom. Boston: Houghton Mifflin Company.

Tripp, A., & Tollefson, N. (1985). Are complex multiple-choice options more difficult and discriminating than conventional multiple-choice options. Journal of Nursing Education.

Weiten, W. (1982). Relative effectiveness of single and double multiple-choice questions in educational measurement. Journal of Educational Research, 5, 46-50.

Wesman, A. G., & Bennett, G. K. (1946). The use of 'none of these' as an option in test construction. Journal of Educational Psychology, 27, 541-549.

Williamson, M. L., & Hopkins, K. D. (1967). The use of 'none of these' versus homogeneous alternatives on multiple-choice tests: Experimental reliability and validity comparisons. Journal of Educational Measurement, 4, 53-57.

Figure 1 Sample Experimental Items

Item Stem — Given Y = 1.5X + 6.0, what is the predicted score on Y for a predictor score of 26?

One-Correct Answer

a. 7.5
b. 39.0
c. 45.0
d. 58.0

"None of the Above" as a foil

a. 7.5
b. 39.0
c. 45.0
d. none of the above

"None of the Above" as the correct response

a. 7.5
b. 39.0
c. 58.0
d. none of the above

Item Stem — George earned a z-score of 0 on a standardized achievement test. Which of the following interpretations of this score is correct?

One-correct Answer

a. George scored at the 50th percentile.
b. George received the lowest score in the class.
c. George answered none of the test items correctly.
d. A mistake was made in scoring George's test.

"None of the Above" as a foil

a. George scored at the 50th percentile.
b. George answered none of the test items correctly.
c. A mistake was made in scoring George's test.
d. none of the above

None of the Above" as the correct response

a. George received the lowest score in the class.
b. George answered none of the test items correctly.
c. A mistake was made in scoring George's test.
d. none of the above

Table 1

Means and Standard Deviations for Item Difficulty and Item Discrimination for Sets of Experimental Items Using Different Formats

Item Set	Item Difficulty		Item Discrimination	
	$\bar{X}$	s	$\bar{X}$	s
E-1 "None of the above" is the correct response	.69	.21	.36	.14
E-2 "None of the above" is a foil	.75	.20	.13	.13
E-3 One-correct answer	.82	.16	.18	.16
C-1 "Correct answer is not given" (computation problems)	.68	.13	.27	.16

Table 2

Means and Standard Deviations for Item Difficulty and Item Discrimination for Set of Nonexperimental Items Using Different Formats

Item Set	Item Difficulty		Item Discrimination	
	$\bar{X}$	s	$\bar{X}$	s
One-correct answer	.72	.18	.50	.14
Complex alternative as the correct answer	.59	.22	.27	.18